Improving the accessibility of digital documents for visually impaired users : Contributions of the Textual Architecture Model

Laurent Sorin, Mustapha Mojahid, Nathalie Aussenac-Gilles, Julie Lemarié

Improving the Accessibility of Digital Documents for Blind Users: Contributions of the Textual Architecture Model

Laurent Sorin[1], Mustapha Mojahid[1], Nathalie Aussenac-Gilles[1], and Julie Lemarié[2]

[1] Institut de Recherche en Informatique de Toulouse, Toulouse, France
{sorin,mojahid,aussenac}@irit.fr
[2] Laboratoire Cognition-Langues-Langage-Ergonomie, Toulouse, France
lemarie@univ-tlse2.fr

Abstract. This paper presents a framework which aims at describing text formatting, based on a model coming from the field of logic and linguistics, the Textual Architecture Model [23]. The goal is to improve documents accessibility for blind users. The project will later focus on evaluating the efficiency of different navigation and content presentation strategies, based on this framework.

Keywords: blind people, documents accessibility, visual signals, text structure.

1 Introduction

Accessibility of information contained in digital documents is a crucial challenge for visually impaired people, especially for blind users. Indeed, it is predicted that the number of blind people will drastically increase with the global ageing of world population. Besides, blind users should be in the center of design issues since Internet and new technologies are an unprecedented opportunity for them to perform tasks that they can hardly do without [8]. When figuring out how to give blind users access to information contained in digital documents, two general approaches may be considered.

The first approach is to change the environment; that is to say creating a new version of each document, media or web page that is entirely and specifically designed for blind users. This is what transcription companies do by offering for instance audio or Braille versions of books. However, this strategy is very rarely used because it is time-consuming and expensive; as far as digital documents are concerned, the always growing body of web pages and digital documents makes this approach impossible to generalize, and thus marginal.

The second possible solution is to include in the original document annotations and possibly additional information in order to make it accessible through assistive technologies. Instead of creating an alternate version of a document, the document designer will respect for instance the WAI accessibility guidelines for visually impaired people (e.g. to give each graphic presentation a textual counterpart), so that a blind user may easily access the information contained in the document.

Even though there has been much effort on designing assistive technologies and accessible information, the situation often remains frustrating for blind users (e.g. [10, 18]). According to [20], several reasons may explain why Internet is still not accessible to blind users. Among them, one should stress the fact that some types of contents are intrinsically visual and can hardly be represented using non-visual modalities. Another major reason is that Internet has a growing multi-task, multi-objects and multi-application logic which makes it particularly difficult for blind users to build a representation of the page layout.

Indeed, digital documents are primarily designed to be visually displayed, so that the expressive means offered by a spatial layout are often intensively used to create complex objects like tables, graphs, outlines, menus, etc. Consequently, it is very challenging to create adaptable contents, i.e. contents that can be presented in different ways without losing structural information and associated cognitive functions. Up to now, when a blind user accesses a document via an audio or a tactile device, he/she has very few clues about the original layout. For instance, Text-To-Speech software (TTS) tools are efficient to oralize a page of continuous text but still struggle with typical web pages [7] and text objects like headings [13].

Yet, many research works in educational and cognitive psychology have clearly shown the positive effects of text signaling devices revealing the text architecture (titles, headings, lists, overviews, etc.) on text processing (see [11] for a review). For instance, a recent study by [13] shows that it is possible to improve oralized texts comprehension by systematically rendering the information conveyed by text headings. In the first experiment, the effects of headings and preview sentences on outlining performances were compared to a control condition (no signal) for both a printed text version and an auditory TTS presentation. It was found that the task of reporting the text organizational structure (outlining task) is facilitated by preview sentences as compared to a no signals condition for both printed text and TTS audio rendering of the same text. Because of their entirely discursive nature, preview sentences are adaptable to audio presentation. In contrast, adding headings to the text was efficient for the printed text but poor for the audio presentation since speech synthesis can't communicate nonverbal information carried by headings in their visual form. In the second experiment, it was further investigated how headings could be rendered by TTS presentation using the analysis provided by [12] on the different information functions associated to headings. Prosodic cues like pauses and discursive indications were added to enrich the TTS presentation and mimic the printed text. The result was that outlining performance improved to levels similar to the visual headings condition of Experiment 1. This shows that giving access to information carried by visual signals in audio format can improve content comprehension.

In this context, the MathArchiTact project aims at allowing blind users to access a document's visual properties and logical structure and at designing new reading tools. We focus on mathematics text-books and try to improve their accessibility to blind high-school teenagers. The project is currently in its early stages.

2 Textual Architecture Model

2.1 Overview

To describe the visual properties of a document, we will use and enrich a model coming from the fields of logic and linguistics: the *Textual Architecture Model* (TAM) [16, 17, 22, 23]. TAM has been used in a computer science perspective for text-generation [14] and analysis [11, 21] and in a psycholinguistic perspective for modeling and predicting the effects of text signaling devices on cognitive processing [11]. Here, as a first step in the MathArchiTact project, we want to use this model to make text formatting available to blind people. Though the scope of this paper is to analyze the model contributions to our goal and examine how to implement it, the longer-term goal is to improve document navigation and content comprehension of documents described with TAM.

This model aims at providing a semantic analysis of text formatting properties that contribute to the "text architecture.

2.2 Key Concepts

According to TAM, a text is composed of (1) a **message** and (2) its specific **formatting**, those two components being separated in the model. A good analogy to understand is to compare it with the data/data-presentation separation, in computer science.

(1)The **message** is made up of the *content* the author intends to communicate to the reader (ideas and concepts), and their specific *linguistic expression* by the author's choice of wording, syntax, and so on, to convey the content. (2)The **text-formatting** properties refer to typography and disposition of the textual content. Those visually distinctive aspects of the text are called **textual objects**; in other words textual objects refer to text-formatting phenomena. For instance: headings, lists, paragraph structure, and any other visually identifiable entity in a text are textual objects.

Finally, the **text architecture** refers to the document textual objects and the relationships that exist between them. The underlying idea here is that text-formatting reflects the author's intention to organize and structure his message; this is why we speak of "architecture".

The text-formatting is represented in TAM as **metatext**, where the metatext is a coherent and cohesive set of metasentences with a specific grammar [17], each metasentence describing the intention that underlies the use of the corresponding text-formatting property. We use the term "**metasentence**", borrowed from Harris [9], to designate language used to describe language itself and its properties (here the text visual properties), as opposed to language referring to real world elements. For instance, "This article is divided into three parts" is a metasentence because it conveys information about the text rather than objects or events in the world.

Fig. 1 shows an example of how text-formatting can be described using metasentences with the TAM.

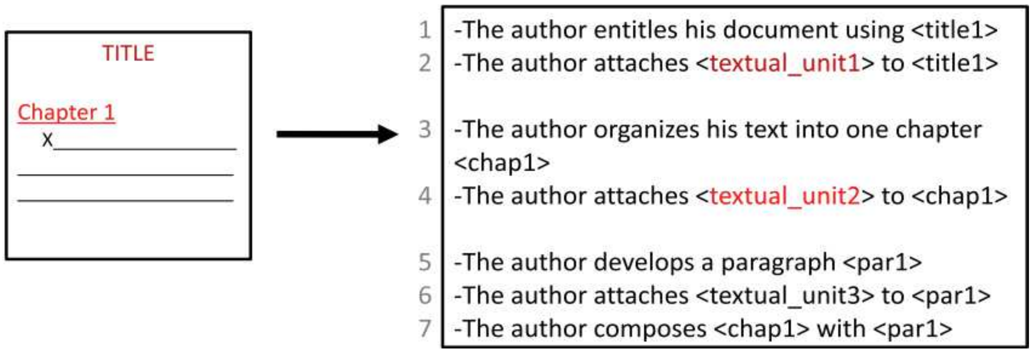


Fig. 1. Example of text-formatting representation using the TAM

As you can see, each textual object has a unique id in the metatext (title1, chap par1…). Metasentences 2, 4 and 5 shows how content is separated from textual o jects, by attaching textual content (divided in textual units) to the corresponding te tual objects. Notice that a textual-object may itself contain different textual objec here the chapter contains a paragraph.

Concerning the relationships between textual objects, two categories appear. T first comprises all of the composition relationships which give the hierarchical stru ture of the text, as in the metasentence 3 in Fig. 1; they depict the text logical stru ture. The other category regroups all of the non linear relationships between objec for instance a footnote annotating a paragraph.

Lastly, even though the previous example included only organizational textual o jects for illustration purposes, the TAM allows describing very local formatti phenomena such as emphasizes.

2.3 Metasentences Properties

Metasentences are the key concept of the model and have several interesting prope ties that are worth being reported.

First, metasentences are generic: one metasentence can describe adequately o class of textual objects. For instance every possible “first level title” textual obje can be described with the same metasentence regardless of their visual for considering a given content for this object.

Besides, a metasentence does not only describe one textual object, it constitutes discursive form. In this way, a metasentence can be “reduced” to its visual form, t corresponding textual object. The concept of reducibility borrows from Harris’s [proposal that metalanguage can be truncated from a complete sentence to a reduc sentence but that the reduction leaves traces in the utterance. Thus the property reducibility is not specific to the realm of texts but is a general property of languag As a consequence, a metasentence may take various forms in the context of the TAM

At one extreme, a metasentence may be left intact, appearing as a discursive stat ment in the text. At the other extreme, a metasentence may be reduced to the po where it is represented only by visual contrast in the text. The example belo illustrates the concept of reducibility.

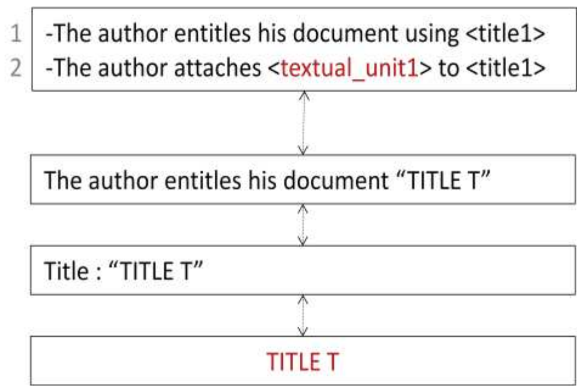


Fig. 2. Reduction / development of a set of metasentences

Fig. 2 shows how the original text at the bottom finds its discursive equivalent in the set of metasentences (in the first frame at the top). The traces left by the reduction process include syntactic transformations, lexical elements (e.g., the lexical content of a specific heading), typographical and spatial realization properties (e.g., italics, bold characters, indentation, blanks), and punctuation marks.

2.4 Technical Framework

After presenting the TAM, the next step in order to apply it to digital documents accessibility is to formally adapt this model, originally a linguistics model, to build a framework for accessing text-formatting semantics as described in the TAM.

Regardless of the original document format, segmentation of textual objects implies being able to annotate the content. Each textual object has its properties and potentially relationships with other objects. Those constraints make markup languages appropriate for implementing the TAM. This choice was also motivated by the wish to be compliant with the DAISY[1] standard which is widely use in digital libraries [6]. Most of DAISY audio-books are in HTML or XML format. The idea would be to keep their existing structure and annotate a duplicated file with the TAM, using the same html/xml ids, which could allow further navigation into textual objects.

For the moment, the annotation process is manual. Yet, depending on the original document format, semi-automatic methods could be proposed. For instance, in the case of web pages, we could use styles defined in CSS[2] to first segment textual objects. More generally, documents where formatting styles are well defined are easier to segment with the TAM. However, a reliable and fully automatic method to segment textual objects appears to be almost impossible since formatting isn't always consistent in one document. The most efficient approach would be to include TAM annotation in the documents production process.

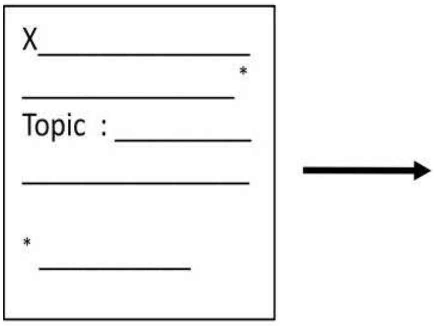
The core of the TAM is the metasentences describing textual objects. Using XML or XHTML, the tags must keep the metasentences following properties: content/formatting

---

[1] International standard for audio books.

[2] Cascading Style Sheets: language used to describe the formatting of several markup languages.

ɔn, using unique ids for objects, include objects properties and describe relation-
tween objects (composition and non-linear relationships, see 2.2).




```
<p id="par_01">
    ...Paragraph content...
    <span class="definition" id="def_01">
        <span class="definition_label">
            Topic
        </span>
        ...Definition Content...
    </span>
</p>
<span class="annotation" id="ann_01">... Annotation...</span>
...
<rel label="annotate" from="ann_01" to="par_01" />
```

Fig. 3. TAM implementation in XML

ent/formatting separation is native with markup languages. Composition rela-s emerge with the tags hierarchy and, for the headings, their order. Textual ids and properties are described with tags properties. Finally we use special l objects ids to note non-linear relationships as the annotation in the example

ntly, the DTD[3] is still under construction and should describe objects proper- relationships. For instance, in Fig. 3, the definition is tagged using the class of the HTML tag “span”, but could also be described using a new tag. A set twenty textual objects and relationships will be formalized; yet new textual can be encountered depending on the studied corpus. For instance with ma-s we can add theorems, lemma, demonstration, etc.
erning the reduction/development properties of metasentences (see 2.3), dis-equivalents are stored in a separate file for each textual object define, for later tion purposes.

Contributions to Digital Documents Accessibility

tion presents how text formatting information made available with the TAM ribute to accessibility of digital documents.

ccessing Content and Structure

ined in 2.3, the TAM makes available discursive counterparts to text objects, from fully discursive metasentences to more reduced forms. For each type of bject of the model we store the different more or less discursive forms of the n order to fit different presentation needs. For instance: “The definition of A is d also be presented with the sentence “A, B” (using prosodic cues). Those two

3 Document Type Definition: tags definition, here for an XML file.

forms of the same object could each be useful depending on the user's task and the specific text object to render.

Currently, several approaches in document presentation for the blind try to give an overview of the content using keywords [1] or summaries [5]. This may be very useful for the user to build a first anchoring representation of the document content and structure, but accessing the visual structure of the document during reading is also crucial.

3.2 Document Navigation

DAISY readers allow basically two types of navigation: local and global navigation[4]. Local navigation refers to text reading control and includes actions such as jumping to the next paragraph or increasing the reading rate, whereas global navigation refers to navigation between text sections and headings. By segmenting documents with the TAM we could allow a new type of navigation through the textual objects relationships. Composition relationships would complete what DAISY call "global navigation" and non-linear relationship would give information about the text structure on a lower level.

Segmenting documents textual objects would also allow to reorganize the content regarding the relationships between objects. For instance, when reading a document, we could regroup textual objects according to their non-linear relationships to avoid cross references, e.g. a theorem with its demonstration and related annotations.

3.3 The case of Mathematics

In the context of the MathArchiTact project, we chose to work on a corpus of mathematics text-books to improve their accessibility for blind high-school teenagers. This choice was driven by the fact that mathematics courses have a very rich visual structure, richer than other courses. Besides, an extensive literature exist on complex mathematics objects access by blind people, mainly formulas [2, 3, 19], which could give us clues about content presentation and navigation.

4 Future Work and Perspectives

We proposed a framework for implementing the TAM in digital documents, in the perspective of their access by blind users.

The next step in the project will be to develop a reader with navigation and presentation techniques adapted to documents annotated with the TAM. Results obtained with this reader will be compared to existing software and methods. This reader will most likely include existing presentation conventions such as prosodic cues, earcons[5]

---

[4] See G. Kerscher: Theory behind the DTbook DTD, `http://data.daisy.org/publications/docs`

[5] Non-verbal sounds used to represent a specific event or convey other information.

and spearcons[6] which are often used in audio interfaces [4, 24]. It could also make use of a Braille output, depending on the user task and presented on textual objects.

Several ways could be explored in the later stages of the project: automatic segmentation of textual objects will be studied, as well as segmentation of the rhetorical structure of the text using the Rhetorical Structure Theory [15] combined with the TAM (merging the two models was already studied for text generation [14]). Thanks to the flexibility offered by markup languages, exporting metadata using web standards would be easy in the case of existing data model like ”Learning object Metadata”[7], and constitute another interesting trail.

Finally, we should mention that the proposed framework could be used for other types of disabilities, such as cognitive impairments, or specific situations where the visual modality is not available (no screen, small screens, etc.) as it provides access to the logical structure of the content.

Acknowledgements. We would like to thank “l’Institut des Jeunes Aveugles de Toulouse” for their precious support in the MathArchiTact project. This work is funded by the “Region Midi-Pyrénées” and the “Pôle de Recherche et d’Enseignement Supérieur de Toulouse”.

References

1. Ahmed, F., Borodin, Y., Puzis, Y., Ramakrishnan, I.: Why Read if You Can Skim: Towards Enabling Faster Screen Reading. In: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A 2012), p. 39 (2012)
2. Alajarmeh, N.: Doing Math: Mathematics Accessibility Issues. In: Proceedings of the Interna-tional Cross-Disciplinary Conference on Web Accessibility (W4A 2012), p. 23. ACM (2012)
3. Awdé, A.: Techniques d’interaction multimodales pour l’accès aux mathématiques par des personnes non-voyantes. PhD Theses, University of Quebec (2009)
4. Bates, E., Fitzpatrick, D.: Spoken Mathematics Using Prosody, Earcons and Spearcons. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 407–414. Springer, Heidelberg (2010)
5. Berger, A.L., Mittal, V.O.: OCELOT: a system for summarizing Web pages. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 144–151. ACM, New York (2000)
6. Christensen, L.B., Stevns, T.: Biblus – A digital library to support integration of visually impaired in mainstream education. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part I. LNCS, vol. 7382, pp. 36–42. Springer, Heidelberg (2012)
7. Fellbaum, K., Kouroupetroglou, G.: Principles of electronic speech processing with applications for people with disabilities. Technology and Disability 20(2), 55–85 (2008)
8. Giraud, S., et al.: L’accessibilité des interfaces informatiques pour les déficients visuels. In: Dinet, J., Bastien, C. (eds.) L’ergonomie des objets et environnements physiques et numériques, Hermes - Sciences Lavoisier, Paris (2011)

---

[6] Spoken audio samples speeded-up at very fast rates, which cannot be recognized as speech when listening.

[7] Data model used to describe a learning object.

9. Harris, Z.S.: Structures mathématiques du langage. Dunod (1968)
10. Lazar, J., Allen, A., Kleinman, J., Malarkey, C.: What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users. International Journal of Human-Computer Interaction 22(3), 247–269 (2007)
11. Lemarié, J., Lorch, R., Eyrolle, H., Virbel, J.: SARA: A Text-Based and Reader-Based Theory of Signaling. Educational Psychologist 43(1), 27–48 (2008)
12. Lemarié, J., Lorch, R., Pery-Woodley, M.-P.: Understanding How Headings Influence Text Processing. Discours 10 (2012), doi:10.4000/discours.8600
13. Lorch, R.F., Chen, H.-T., Lemarie, J.: Communicating headings and preview sentences in text and speech. Journal of Experimental Psychology: Applied 18(3), 265–276 (2012)
14. Luc, C.: Représentation et composition des structures visuelles et rhétoriques du texte. Approche pour la génération de textes formatés (2000)
15. Mann, W.C., Thompson, S.A.: Rhetorical Structure theory description and constructions of text structures. In: Kempen, G. (ed.) Natural Language Generation. NATO ASI Series, Series E: Applied Sciences, vol. 135, pp. 85–96. Martinus Nijhoff Publishers, Dordrecht (1987)
16. Pascual, E.: Integrating Text Formatting and Text Generation. Trends in Natural Language Generation: An Artificial Intelligence Perspective, pp. 205–221 (1996)
17. Pascual, E., Virbel, J.: Semantic and Layout Properties of Text Punctuation, 41–48 (1991)
18. Petit, G., Dufresne, A., Robert, J.-M.: Introducing TactoWeb: A Tool to Spatially Explore Web Pages for Users with Visual Impairment. In: Stephanidis, C. (ed.) Universal Access in HCI, Part I, HCII 2011. LNCS, vol. 6765, pp. 276–284. Springer, Heidelberg (2011)
19. Schweikhardt, W., Bernareggi, C., Jessel, N., Encelle, B., Gut, M.: LAMBDA: A European system to access mathematics with Braille and audio synthesis. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2006. LNCS, vol. 4061, pp. 1223–1230. Springer, Heidelberg (2006)
20. Uzan, G., Wagtaf, P.: A Model and Methods to solve problems of Accessibility and Information for the Visually Impaired. In: Proceedings of the International Conference STHESCA, Krakow (2011)
21. Vigouroux, N., et al.: Problématique, enjeux et perspectives de la présentation orale de documents: une approche pluridisciplinaire. In: Inscription Spatiale du Langage structure et processus (ISLsp 2002), Toulouse, January 29-30, pp. 139–150. IRIT (2002)
22. Virbel, J.: Langage et métalangage dans le texte du point de vue de l'édition en informatique textuelle. Cahiers de grammaire. pp. 5–72 (1985)
23. Virbel, J.: Structured documents. In: André, J., et al. (eds.) Structured Documents, pp. 161–180. Cambridge University Press, NY (1989)
24. Walker, B.N., Nance, A., Lindsay, J.: SPEARCONS: Speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the 12th International Conference on Auditory Display, pp. 63–68 (2006)